*Sankhyā : The Indian Journal of Statistics*
1979, Volume 40, Series B, Pts. 3 and 4, pp. 204–208.

# ON ROBUSTNESS OF DESIGNS AGAINST INCOMPLETE DATA

*By* SUBIR GHOSH
*Indian Statistical Institute*

*SUMMARY.* In this paper, we characterize the robustness property of designs against incomplete data in the sense that, when any $t$ (a positive integer) observations are missing, all parameters are still estimable in the model assumed. We also present some examples of Srivastava-Chopra Optimum balanced resolution $V$ plans for $2^m$ factorials which are robust against missing of any two observations.

## 1. Introduction

The robustness of designs against incomplete data in case of missing of any single observation was first considered in Ghosh (1978). This paper gives a characterization of robustness property in the general case of missing of any $t$ observations. Some examples of designs robust against missing of any two observations are also presented.

## 2. Robust designs

Consider the ordinary linear model

$$E(\boldsymbol{y}) = A\boldsymbol{\xi} \qquad \dots \quad (1)$$

$$V(\boldsymbol{y}) = \sigma^2 I_N \qquad \dots \quad (2)$$

$$\text{Rank } \boldsymbol{A} = \nu \qquad \dots \quad (3)$$

where $\boldsymbol{y}(N\times 1)$ is a vector of observations, $\boldsymbol{A}(N\times\nu)$ is a known matrix, $\boldsymbol{\xi}(\nu\times 1)$ is a vector of fixed unknown parameters and $\sigma^2$ is a constant which may or may not be known. Let $T$ be the underlying design corresponding to $\boldsymbol{y}$.

*Definition* 1: A design under the model (1–3) is said to be robust against missing of any $t$ (a positive integer) observations if the $(N-t\times\nu)$ matrix obtained from $\boldsymbol{A}$ by omitting any $t$ rows has rank $\nu$. It is clear from definition 1 that $N$ must at least be $\nu+t$. Suppose $N = \nu+k$, where $k(\geqslant t)$ a positive integer. Clearly, there exist $k$ linearly independent vectors $\boldsymbol{C}_i' = (C_{i1}, \dots, C_{iN})$, $i = 1, \dots, k$, with real elements satisfying

$$\boldsymbol{C}_i'A = \boldsymbol{0} \qquad \dots \quad (4)$$

Consider the $(k \times N)$ matrix

$$C = \begin{bmatrix} C_{11} & C_{12} & \ldots & C_{1t} & \ldots & C_{1N} \\ C_{21} & C_{22} & \ldots & C_{2t} & \ldots & C_{2N} \\ C_{k1} & C_{k2} & \ldots & C_{kt} & \ldots & C_{kN} \end{bmatrix} \qquad \ldots \quad (5)$$

whose $i$-th row is $\boldsymbol{C}_i'$ and furthermore, Rank $\boldsymbol{C} = k$. We now recall that a matrix $\boldsymbol{B}$ is said to have the property $P_t$ if no $t$ columns of $B$ are linearly dependent. The following theorem characterizes the robustness property.

Theorem 1 : *Let $T$ be a design under* (1-3) *with $N = \nu + k$ observations, where $k(\geqslant t)$ a positive integer. Then, $T$ is robust against missing of any $t$ observations if and only if (iff) the matrix $\boldsymbol{C}$, defined in* (5), *has the property $P_t$.*

*Proof :* Suppose $\boldsymbol{C}$ has $P_t$. Let

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_1 \\ \ldots \\ \boldsymbol{A}_2 \end{bmatrix}, \quad \boldsymbol{C} = [\boldsymbol{C}_1^* : \boldsymbol{C}_2^*], \qquad \ldots \quad (6)$$

where $\boldsymbol{A}_1(t \times \nu)$, $\boldsymbol{A}_2(\overline{N-t} \times \nu)$, $\boldsymbol{C}_1^*(k \times t)$ and $\boldsymbol{C}_2^*(k \times \overline{N-t})$.
We have, from (4),

$$\boldsymbol{C}_1^* \boldsymbol{A}_1 + \boldsymbol{C}_2^* \boldsymbol{A}_2 = 0. \qquad \ldots \quad (7)$$

Suppose

$$\boldsymbol{C}_1^* = \begin{bmatrix} \boldsymbol{C}_{11}^* \\ \boldsymbol{C}_{12}^* \end{bmatrix}, \quad \boldsymbol{C}_2^* = \begin{bmatrix} \boldsymbol{C}_{21}^* \\ \boldsymbol{C}_{22}^* \end{bmatrix} \qquad \ldots \quad (8)$$

where $\boldsymbol{C}_{11}^*(t \times t)$, $\boldsymbol{C}_{12}^*(\overline{k-t} \times t)$, $\boldsymbol{C}_{21}^*(t \times \overline{N-t})$, and $\boldsymbol{C}_{22}^*(\overline{k-t} \times \overline{N-t})$. Suppose, furthermore, Rank $(\boldsymbol{C}_{11}^*) = t$. Thus, we get

$$\boldsymbol{C}_{11}^* \boldsymbol{A}_1 + \boldsymbol{C}_{21}^* \boldsymbol{A}_2 = 0. \qquad \ldots \quad (9)$$

Hence,

$$\boldsymbol{A}_1 = -\boldsymbol{C}_{11}^{*-1} \boldsymbol{C}_{21}^* \boldsymbol{A}_2. \qquad \ldots \quad (10)$$

Thus, the rows of $\boldsymbol{A}_1$ are linear combinations of the rows in $\boldsymbol{A}_2$. Therefore, the matrix $\boldsymbol{A}_2$ obtained from $\boldsymbol{A}$ by omitting $t$ rows in $\boldsymbol{A}_1$, has rank $\nu$. The argument is similar for any other set of $t$ rows of $\boldsymbol{A}$. Hence the design $T$ is robust.

Suppose the design $T$ is robust against missing of $t$ observations. Then, there is a $(t \times \overline{N-t})$ matrix $\boldsymbol{D}$ satisfying

$$\boldsymbol{A}_1 = \boldsymbol{D}\boldsymbol{A}_2 \qquad \ldots \quad (11)$$

i. e.,

$$[\boldsymbol{I}_t : -\boldsymbol{D}] \begin{bmatrix} \boldsymbol{A}_1 \\ \boldsymbol{A}_2 \end{bmatrix} = 0. \qquad \ldots \quad (12)$$

Considering (4), (6), and (9), it follows that there exists a $(t \times k)$ matrix $\boldsymbol{U}$ such that

$$\boldsymbol{U}\boldsymbol{C}_1^* = \boldsymbol{I}_t, \quad \boldsymbol{U}\boldsymbol{C}_2^* = -\boldsymbol{D}. \qquad \ldots \quad (13)$$

It is now easy to check that Rank $(\boldsymbol{C}_1^*) = t$. Therefore $\boldsymbol{C}$ has $P_t$. This completes the proof of the theorem.

The following results are of practical importance.

Corollary 1: *Suppose $t = 1$. The design $T$ is robust against missing of any one observation iff $\boldsymbol{C}(k \times N)$ has the property $P_1$ or, in other words,*

$$(C_{1j}, C_{2j}, \ldots, C_{kj}) \neq (0, 0, \ldots, 0) \textit{ for } (j = 1, \ldots, N)$$

(i.e., none of the column vectors in $C$ is a null vector).

Corollary 2 : *Suppose $t = 2$. The design $T$ is robust against missing of any two observations iff $\boldsymbol{C}(k \times N)$ has the property $P_2$, or in other words,*

(i) $(C_{1j}, C_{2j}, \ldots, C_{kj}) \neq (0, 0, \ldots, 0)$ *for* $(j = 1, \ldots, N)$,

(ii) $(C_{1j}, C_{2j}, \ldots, C_{kj}) \neq w(C_{1j'}, C_{2j'}, \ldots, C_{kj'})$,

*where $j \neq j'$, $(j, j' = 1, \ldots, N)$, and $w$ is a real constant.*

It is to be remarked that the above results are also true in case $\boldsymbol{A}(N \times M)$, $\boldsymbol{\xi}(M \times 1)$ and Rank $(\boldsymbol{A}) = \nu < \min(M, N)$.

## 3. Examples from $2^m$ factorials

Consider a $2^m$ factorial experiment. The treatments are denoted by $(x_1, x_2, \ldots, x_m)$, where $x_i = 0$ or 1. We denote a design with $N$ treatments by a $(N \times m)$ matrix $\boldsymbol{T}$ whose rows are treatments. Optimal balanced resolution $V$ plans for $2^m$ factorials, $4 \leqslant m \leqslant 8$, and for practical values of $N$, have been presented in the papers of Srivastava and/or Chopra.

By 'weight' of a vector, we mean the number of nonzero elements in it. Let $S_i$ be the set of all $(1 \times m)$ vectors, with elements 0 and 1, of weight $i$ $(i = 0, 1, \ldots, m)$. Clearly the number of members in $S_i$ is $\binom{m}{i}$.

Srivastava-Chopra designs are denoted by $\boldsymbol{\lambda}' = (\lambda_0, \lambda_1, \ldots, \lambda_m)$ where $\lambda_i$ is the number of times the set $S_i$ occurs in the design. Thus $N = \sum_{i=0}^{m} \binom{m}{i} \lambda_i$. These optimum designs may or may not remain optimum or even resolution $V$ plans when some observations are missing. We now present, as example, designs which are robust against missing of any $t$ observations. These designs remain as resolution $V$ plans when any $t$ observations are missing.

*Example* 1 : Consider $m = 4, N = 15$. Here, $\nu = 11$. Thus $k = 4$. The design is represented as $\boldsymbol{\lambda}' = (1\ 1\ 1\ 1\ 0)$. The matrix $\boldsymbol{C}$ is given below

$$\boldsymbol{C} = \begin{bmatrix} 1 & -3 & -3 & -3 & -3 & 2 & 2 & 2 & 2 & 2 & 2 & -1 & -1 & -1 & -1 \\ 0 & 1 & -1 & 1 & -1 & 0 & -2 & 0 & 0 & 2 & 0 & 1 & -1 & 1 & -1 \\ 0 & 1 & -1 & -1 & 1 & 0 & 0 & -2 & 2 & 0 & 0 & -1 & 1 & 1 & -1 \\ 0 & 1 & 1 & 1 & -1 & -2 & 0 & 0 & 0 & 0 & 2 & 1 & 1 & -1 & -1 \end{bmatrix}$$

It is easy to check that the above matrix has the property $P_2$ but not $P_3$. Thus the present design is robust against missing of any two observations and not robust against missing of any three observations. Clearly for $N = 16$, the design $\boldsymbol{\lambda}' = (1\ 1\ 1\ 1\ 1)$ is also robust against missing of two observations.

*Example* 2 : Consider $m = 5$, $N = 22a$. We have $\nu = 16$ and thus $(k = 6)$. The design is given by $\boldsymbol{\lambda}' = (1\ 1\ 1\ 0\ 1\ 1)$. We present the matrix $\boldsymbol{C}$ below

$$\boldsymbol{C} = \begin{bmatrix} 3 & -3 & -3 & -3 & -3 & -3 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & -3 & -3 & -3 & -3 & -3 & 7 \\ -3 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & -1 & -1 & -1 & 3 \\ 0 & -2 & 2 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & -1 & -1 & -1 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & -2 & 0 & 2 & 0 & 0 & 1 & 0 & 1 & 1 & -1 & 0 & 0 & -1 & -1 & 0 & 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & -2 & 0 & 0 & 2 & 0 & 1 & 1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & -1 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & -2 & 0 & 0 & 0 & 2 & 1 & 1 & 1 & 0 & 0 & 0 & -1 & 0 & -1 & -1 & 1 & 0 & 0 & 0 & -1 & 0 \end{bmatrix}$$

Observe that the above matrix has the property $P_2$ and, therefore, this design is robust against missing of any two observations. It is clear that the designs $N = 23a$, $\boldsymbol{\lambda}' = (2\ 1\ 1\ 0\ 1\ 1)$, $N = 24a$, $\boldsymbol{\lambda}' = (2\ 1\ 1\ 0\ 1\ 2)$, and $N = 25a, \boldsymbol{\lambda}' = (3\ 1\ 1\ 0\ 1\ 2)$ have the same property.

## References

Bose, R. C. (1947): Mathematical theory of symmetrical factorial designs. *Sankhyā*, **8**, 107-66.

Box, G. E. P. and Draper, N. R. (1975): Robust designs. *Biometrika*, **62**, 2, 347-52.

Ghosh, S. (1978): On robustness of optimal balanced resolution $V$ plans. (Submitted for publication)

Herzberg, A. M. and Andrews, D. F. (1976): Some considerations in the optimal design of experiments in non-optimal situations. *J. Roy. Stat. Soc.* Ser. B, **3**, 284-89.

Huber, P. (1975): *Robustness and Designs. A Survey of Statistical Design and Linear Models* (Ed: J. N. Srivastava), Amsterdam: North-Holland Publishing Company, 287-301.

Srivastava, J. N. and Chopra, D. V. (1971): Balanced optimal $2^m$ fractional factorial designs of resolution $V$, $m < 6$. *Technometrics*, **13**, 257-69.

*Paper received: October, 1978.*
*Revised: January, 1979.*